

Performance Evaluation of Transformer Models: Scratch, Bart, and Bert for News Document Summarization

Daurin Nabilatul Munna¹, Enggarani Wahyu Ekaputri², Khadijah Fahmi Hayati Holle^{*3}

^{1,2,3}Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia

Email: khadijah.holle@uin-malang.ac.id

Received : Jul 18, 2024; Revised : Oct 30, 2024; Accepted : Nov 14, 2024; Published : Apr 26, 2025

Abstract

This study evaluates the performance of three Transformer models: Transformer from Scratch, BART (Bidirectional and Auto-Regressive Transformers), and BERT (Bidirectional Encoder Representations from Transformers) in the task of summarizing news documents. The evaluation results show that BERT excels in understanding the bidirectional context of text, with a ROUGE-1 value of 0.2471, ROUGE-2 of 0.1597, and ROUGE-L of 0.1597. BART shows strong ability in de-noising and producing coherent summaries, with a ROUGE-1 value of 0.5239, ROUGE-2 of 0.3517, and ROUGE-L of 0.3683. Transformer from Scratch, despite requiring large training data and computational resources, produces good performance when trained optimally, with ROUGE-1 scores of 0.7021, ROUGE-2 scores of 0.5652, and ROUGE-L scores of 0.6383. This evaluation provides insight into the strengths and weaknesses of each model in the context of news document summarization.

Keywords : BART, BERT, Document Summarization, NLP, ROUGE, Transformer.

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



1. INTRODUCTION

Model berbasis Transformer telah mendominasi berbagai tugas dalam Natural Language Processing (NLP), termasuk tugas summarization atau peringkasan dokumen. Peringkasan dokumen bertujuan untuk menghasilkan versi singkat dari teks asli dengan mempertahankan informasi penting. Di antara berbagai model yang digunakan untuk tugas ini, Transformer from Scratch, BART (Bidirectional and Auto-Regressive Transformers), dan BERT (Bidirectional Encoder Representations from Transformers) adalah beberapa arsitektur terkemuka yang menunjukkan kinerja unggul.

Transformer from Scratch adalah model yang dibangun dan dilatih dari awal tanpa menggunakan model pralatih[1]. Keuntungan dari pendekatan ini adalah kemampuan untuk menyesuaikan arsitektur model sesuai dengan kebutuhan spesifik, meskipun membutuhkan data pelatihan dalam jumlah besar serta waktu dan sumber daya komputasi yang signifikan[2]. Di sisi lain, BART merupakan model yang menggabungkan pendekatan autoregressive dan auto-encoding, memberikan performa yang kuat dalam berbagai tugas NLP termasuk peringkasan dokumen[3]. BART melakukan denoising pada input teks dan menghasilkan output yang lebih terstruktur dan koheren, menjadikannya efektif dalam tugas peringkasan.

Sementara itu, BERT dikenal karena kemampuannya dalam memahami konteks dua arah dari teks, yang sangat bermanfaat dalam memahami dan meringkas dokumen yang kompleks[4]. Meskipun BERT awalnya dirancang untuk tugas-tugas seperti klasifikasi teks dan named entity recognition, adaptasinya untuk peringkasan dokumen telah menunjukkan hasil yang menjanjikan.

Penelitian di bidang peringkasan dokumen telah mengalami perkembangan pesat dengan munculnya berbagai model berbasis Transformer, seperti BERT (Bidirectional Encoder Representations from Transformers), BART (Bidirectional and Auto-Regressive Transformers), dan T5 (Text-to-Text

Transfer Transformer). BERT, yang diperkenalkan oleh Devlin et al. [5], telah diterapkan dalam berbagai tugas NLP termasuk peringkasan teks, dengan hasil yang menunjukkan peningkatan kinerja dibandingkan metode sebelumnya. Sementara itu, BART, yang dikembangkan oleh Lewis et al. [6], menggabungkan kelebihan dari model encoder-decoder seperti Transformer dan model autoregressive seperti GPT, menunjukkan hasil yang menjanjikan dalam berbagai tugas generasi teks, termasuk peringkasan dokumen. Model T5, yang diperkenalkan oleh Raffel et al. [7], menggunakan pendekatan "text-to-text" di mana semua tugas NLP diformulasikan sebagai masalah transformasi teks ke teks. Implementasi T5 untuk peringkasan dokumen dalam bahasa Indonesia telah dilakukan oleh Purnama dan Utami [8] dengan hasil yang menunjukkan efektivitas model dalam menghasilkan ringkasan yang akurat dan relevan. Selain itu, penelitian terbaru menunjukkan bahwa model hibrida yang menggabungkan berbagai teknik NLP, seperti K-Means clustering dengan model BERT, dapat meningkatkan akurasi peringkasan teks. Raihanunnisa et al. [9] menunjukkan bahwa pendekatan hibrida ini efektif dalam sistem peringkasan teks artikel berita bahasa Inggris. Penelitian yang lebih lanjut oleh Jaiswal et al. [10] mengungkapkan pentingnya sparsitas dalam model pelatihan besar, yang berimplikasi pada optimasi kinerja dan efisiensi model dalam tugas-tugas NLP termasuk peringkasan dokumen.

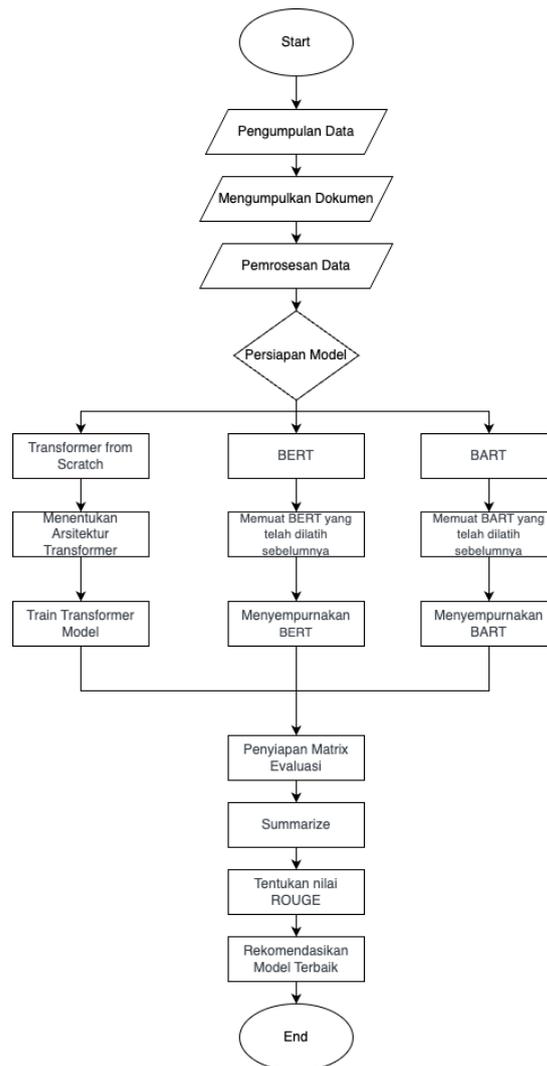
Dalam dunia yang semakin dipenuhi oleh informasi, kemampuan untuk meringkas dokumen secara efisien dan akurat menjadi semakin krusial [11]. Oleh karena itu, penelitian ini tidak hanya memberikan kontribusi pada pengembangan teknologi NLP, tetapi juga berperan dalam meningkatkan efisiensi dan efektivitas dalam pengelolaan informasi [12].

Evaluasi terhadap model Transformer from Scratch, BART, dan BERT dalam konteks peringkasan dokumen menjadi penting untuk mengidentifikasi kekuatan dan kelemahan masing-masing model [13]. Penelitian ini bertujuan untuk memberikan analisis komprehensif mengenai performa ketiga model tersebut dalam tugas peringkasan dokumen, dengan fokus pada efektivitas dan efisiensi setiap model dalam menghasilkan ringkasan yang akurat dan informatif. Efektivitas model dalam meringkas dokumen diukur dari kemampuannya menghasilkan ringkasan yang mencakup informasi penting dengan akurat. Efisiensi model dilihat dari kecepatan dan penggunaan sumber daya yang minimal dalam menghasilkan ringkasan tersebut. [14]. Melalui evaluasi ini, diharapkan dapat diperoleh wawasan yang lebih mendalam tentang perkembangan terbaru dalam teknologi peringkasan dokumen, serta panduan bagi praktisi NLP dalam memilih model yang paling sesuai untuk aplikasi mereka.

Dalam dunia yang semakin dipenuhi oleh informasi, kemampuan untuk meringkas dokumen secara efisien dan akurat menjadi semakin krusial. Natural Language Processing (NLP) sangat berpengaruh dalam meningkatkan efisiensi dan efektivitas berbagai proses analisis data. Dengan kemampuan untuk memahami, menganalisis, dan menghasilkan bahasa alami, NLP dapat mengotomatisasi tugas-tugas yang sebelumnya memerlukan intervensi manusia [15]. Hal ini tidak hanya mengurangi waktu yang dibutuhkan untuk memproses informasi, tetapi juga meningkatkan akurasi dalam pengambilan keputusan berbasis data. Oleh karena itu, penelitian ini tidak hanya memberikan kontribusi pada pengembangan teknologi NLP, tetapi juga berperan dalam meningkatkan efisiensi dan efektivitas dalam pengelolaan informasi.

2. METHOD

Design System adalah alur sistem yang dirancang oleh peneliti untuk melakukan penelitian. Design System yang digunakan digambarkan dalam Gambar 1.



Gambar 1. Alur Design System

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari artikel berbahasa Inggris. Terdapat dua kolom utama dalam dataset ini, yaitu kolom kategori dan kolom teks. Dataset pada penelitian ini tersedia secara publik di platform penyedia data, Kaggle. File yang diunggah adalah sebuah CSV bernama `bbc-text.csv` dapat diakses melalui link berikut <https://www.kaggle.com/datasets/yufengdev/bbc-fulltext-and-category?select=bbc-text.csv>. File ini berisi 2225 baris dan 2 kolom, yang mencakup kategori dan teks lengkap. Kolom `category` menunjukkan kategori dari setiap entri teks dan kolom `text` berisi teks lengkap, semuanya bertipe `object`. Data ini mencakup berbagai kategori seperti teknologi, bisnis, olahraga, dan hiburan, dan disusun untuk digunakan dalam pemrosesan bahasa alami (NLP).

2.2. Preprocessing

Pra-pemrosesan data dalam penelitian ini bertujuan untuk membersihkan dan mempersiapkan data teks agar sesuai untuk analisis lebih lanjut. Dengan melakukan pra-pemrosesan ini dapat meningkatkan kualitas hasil, serta mengurangi noise atau gangguan yang dapat mempengaruhi interpretasi terkait konsistensi, dan memfasilitasi representasi yang lebih baik dari konten teks[16]. Dengan melakukan pra-pemrosesan ini dapat meningkatkan kualitas hasil, serta mengurangi noise atau gangguan yang dapat mempengaruhi interpretasi akhir dari data teks.

2.3. Pembagian Data

Pembagian data dilakukan setelah data yang dikumpulkan dalam proses summarize. Pembagian ini penting karena untuk memastikan bahwa model yang dibuat dapat dilatih dan diuji secara efektif dan objektif. Datasheet akan dibagi menjadi 2 bagian yaitu untuk proses training dan proses testing dengan persentase pada Tabel 1.

Tabel 1. Persentase Pembagian Data

No	Pembagian Data	Presentase
1.	Data Training	80%
2.	Data Testing	20%

2.4. Tokenisasi

Tokenisasi adalah proses membagi teks atau kalimat menjadi unit-unit yang lebih kecil, yang disebut sebagai token. Tujuan dari tokenisasi adalah untuk memecah teks menjadi kata-kata atau unit-unit yang lebih kecil sehingga dapat diolah lebih lanjut. Dengan melakukan tokenisasi, setiap kata dalam teks akan menjadi token tersendiri, yang dapat dihitung frekuensinya, diekstraksi fitur, atau diolah lebih lanjut menggunakan algoritma-algoritma pemrosesan teks[17]. Dalam penelitian ini, terdapat tiga model transformer yang digunakan adalah BERT, BART, dan Transformer from scratch.

Pertama, untuk model BERT, tokenisasi dilakukan dengan menggunakan WordPiece tokenization, dimana teks dibagi menjadi potongan-potongan kecil yang disebut token, termasuk kata-kata dan sub-kata. Proses ini memungkinkan BERT untuk memahami struktur kata yang kompleks dan mengatasi masalah out-of-vocabulary (OOV) dengan menggabungkan kata-kata yang sering muncul menjadi token tunggal, sehingga meningkatkan kinerja model dalam memahami teks[18].

Kedua, dalam model BART, tokenisasi menggunakan Byte-Pair Encoding (BPE), sebuah teknik yang membagi teks menjadi sub-unit yang lebih kecil berdasarkan kemunculan frekuensi. Proses BPE ini memungkinkan BART untuk menangkap makna dari frasa-frasa yang umum dan tidak umum, sehingga memberikan fleksibilitas yang lebih besar dalam pemrosesan teks[19]. Selain itu, BART juga menerapkan tokenisasi khusus untuk tugas summarization, yang memungkinkan model untuk menghasilkan ringkasan yang lebih berkualitas dari teks input.

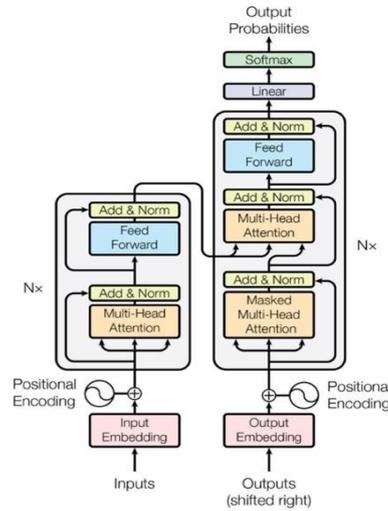
Ketiga, untuk model transformer dari awal (from scratch) dengan tokenisasi menggunakan Spacy, pendekatan yang digunakan adalah menggunakan tokenizer yang dibuat sendiri berdasarkan kamus kata-kata yang ada. Proses tokenisasi ini melibatkan pengenalan kata-kata dalam teks dan pemisahan kata-kata berdasarkan spasi[20]. Meskipun sederhana, pendekatan ini memungkinkan model untuk memahami teks dengan baik, terutama jika kamus yang digunakan kaya dengan variasi kata-kata. Namun, model yang dibangun dari awal memerlukan lebih banyak usaha untuk disesuaikan dengan tugas khusus seperti summarization, tetapi dapat memberikan fleksibilitas yang lebih besar dalam pengembangan model yang dioptimalkan untuk tugas tertentu.

2.5. Penambahan Model

Dalam penelitian ini, terdapat tiga model transformer yang digunakan adalah BERT, BART, dan Transformer from scratch.

Gambar 2 menunjukkan sebuah arsitektur dari model BERT. BERT sendiri merupakan sebuah model yang dilatih untuk memprediksi kata yang hilang dalam konteks secara bidirectional[22]. Proses ini melibatkan beberapa tahapan. Pertama, input urutan teks di-embed menjadi representasi vektor. Selanjutnya, BERT menggunakan lapisan encoder yang sama dengan Transformer standard. Lapisan ini dilengkapi dengan bidirectional self-attention, yang memungkinkan setiap token dalam urutan input untuk memperhatikan setiap token lainnya[24]. Kemudian, terdapat lapisan Feed-Forward Network dan Layer Normalization yang membantu menangkap representasi yang lebih kompleks dari data input.

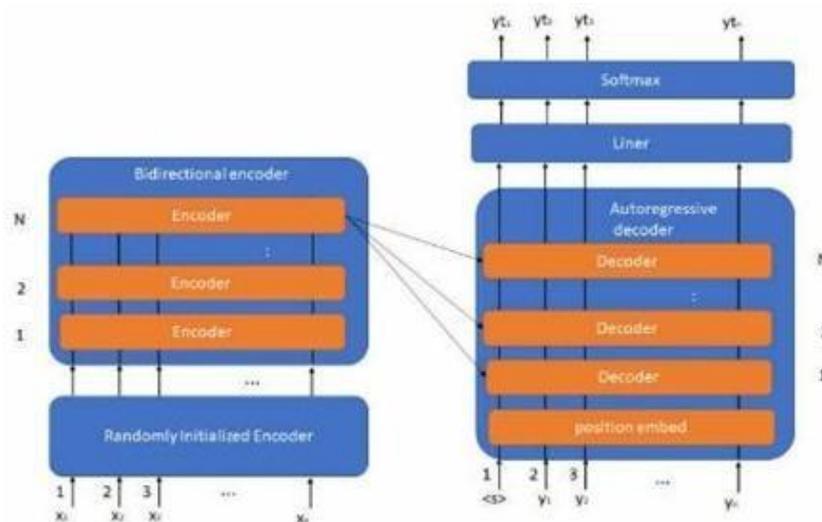
Output dari encoder ini dapat digunakan untuk berbagai tugas NLP, seperti klasifikasi, penjawaban pertanyaan, dan summarization.



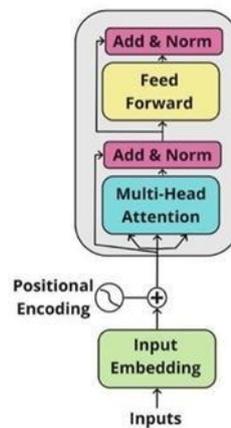
Gambar 2. Arsitektur Model BERT [17]

2.6. Arsitektur Model BART

Gambar 3 menunjukkan sebuah arsitektur dari model BART. BART sendiri adalah sebuah model yang menggabungkan keunggulan dari Transformer encoder dan decoder[25]. Dalam encoder, terdapat stack encoder yang mirip dengan Transformer standar, yang terdiri dari beberapa lapisan. Lapisan-lapisan ini mencakup multi-head attention dan feed-forward networks dengan mekanisme "Add & Norm". Keuntungan utama dari encoder adalah kemampuannya untuk menggunakan konteks bidireksional pada proses encoding. Sementara itu, dalam decoder, terdapat stack decoder yang juga memiliki beberapa lapisan. Decoder ini menggunakan "masked multi-head attention" untuk memastikan bahwa setiap posisi hanya dapat memperhatikan posisi sebelumnya dalam urutan output. Selain itu, decoder juga memiliki lapisan cross-attention yang memperhatikan output dari encoder, serta feed-forward networks dan mekanisme "Add & Norm"[26]. Proses decoding ini menghasilkan probabilitas output melalui lapisan linear dan softmax. Pendekatan ini memungkinkan penggunaan autoregressive pada decoding, yang merupakan salah satu keuntungan utama dari BART.



Gambar 3. Arsitektur Model BART [20]



Gambar 4. Arsitektur Model Transformer from Scratch [22]

Gambar 4 menunjukkan sebuah arsitektur dari model Transformer. Transformer sendiri merupakan sebuah arsitektur model yang menggunakan encoder-decoder stack untuk memproses urutan data[23]. Encoder terdiri dari beberapa lapisan yang mirip, dengan setiap lapisan memiliki dua sub-lapisan utama: Multi-Head Attention dan Feed-Forward Network. Multi-Head Attention memungkinkan model untuk memperhatikan informasi dari berbagai posisi dalam urutan input dengan menggunakan beberapa kepala (heads). Hal ini memungkinkan model untuk menangkap berbagai jenis informasi dari input. Setelah melewati lapisan Multi-Head Attention, inputnya melewati jaringan feedforward yang sama untuk setiap posisi dalam urutan. Setiap sub-lapisan ini diikuti oleh mekanisme "Add & Norm" yang melakukan residual connection (penjumlahan input asli dengan output sub-lapisan) dan normalisasi lapisan. Input embedding dan positional encoding ditambahkan di awal untuk memberikan informasi posisi ke model. Output dari encoder kemudian diteruskan ke blok decoder untuk proses selanjutnya.

2.7. Training Model

Dalam penelitian ini, terdapat tiga model transformer yang digunakan adalah BERT, BART, dan Transformer from scratch. Berikut merupakan penjelasan dalam proses training dari setiap model tersebut:

Pertama, model BART dilatih dengan metode supervised learning yang berfokus pada conditional generation, di mana model diajarkan untuk memprediksi output berdasarkan input yang diberikan. Kelebihan utama BART terletak pada kemampuannya menghasilkan teks yang halus dan konsisten, baik dalam tugas generasi teks maupun penerjemahan. Selain itu, dengan arsitektur transformer-nya, BART mampu menangani input berukuran besar dengan efisien. Namun, proses training BART bisa menjadi lambat karena kompleksitasnya dan memerlukan dataset yang besar serta waktu training yang cukup lama untuk mencapai kinerja yang optimal.

Kedua, model BERT juga menggunakan metode supervised learning, umumnya untuk tugas klasifikasi teks. Dalam penelitian ini, BERT dilatih untuk melakukan klasifikasi urutan teks. Kelebihan utama BERT adalah kemampuannya menghasilkan representasi kata yang kaya dengan memanfaatkan konteks dari seluruh urutan, sehingga cocok untuk berbagai tugas pemrosesan bahasa alami. Selain itu, kemampuannya dalam mengatasi masalah overfitting dengan teknik seperti dropout membuatnya lebih tahan terhadap overfitting pada dataset yang relatif kecil[27]. Namun, BERT memerlukan jumlah parameter yang besar dan sumber daya komputasi yang signifikan, serta waktu training yang bisa menjadi lambat terutama pada dataset yang besar.

Ketiga, model Transformer dari awal dibangun menggunakan Python dan PyTorch, tanpa menggunakan pre-trained weights[28]. Model ini dapat disesuaikan sepenuhnya dengan kebutuhan dan tugas spesifik tanpa bergantung pada pre-trained weights. Proses trainingnya bisa lebih cepat

dibandingkan dengan model yang lebih besar seperti BERT atau BART karena model ini mungkin lebih sederhana. Namun, pembangunan model Transformer dari awal memerlukan pengetahuan yang mendalam tentang arsitektur Transformer dan implementasinya dalam PyTorch, serta memerlukan lebih banyak waktu dan usaha untuk melakukan eksperimen dan penyetelan model agar mencapai kinerja yang optimal.

2.8. Summarization Document

Dalam penelitian ini, terdapat tiga model transformer yang digunakan adalah BERT, BART, dan Transformer from scratch. Berikut merupakan penjelasan dalam merangkum dokumen dalam Bahasa Inggris dari setiap model tersebut :

Pertama, dalam model BERT summarize dilakukan dengan memanfaatkan model yang telah dilatih untuk klasifikasi urutan teks. Fungsi summarize dirancang untuk menerima teks input, yang kemudian diolah oleh tokenizer untuk mengkonversinya menjadi token yang dapat dimengerti oleh model. Setelah itu, teks yang telah di-tokenisasi dimasukkan ke dalam model BERT. Model menghasilkan logits, yang merepresentasikan probabilitas untuk setiap kelas dalam konteks klasifikasi teks. Dengan menggunakan argmax, probabilitas ini dikonversi menjadi label terbaik yang merepresentasikan ringkasan dari teks input, dan kemudian teks ringkasan tersebut didekode menggunakan tokenizer.

Kedua, dalam model BART summarize dilakukan dengan memanfaatkan kemampuan model BART dalam generasi teks. Fungsi print summary menerima artikel sebagai input, kemudian menggunakan model BART untuk menghasilkan ringkasan dari artikel tersebut. Setelah proses generasi selesai, menggunakan pustaka ROUGE, sistem memberikan skor ke akurasi ringkasan yang dihasilkan berdasarkan referensi yang telah disediakan. Proses ini memungkinkan evaluasi objektif terhadap kualitas ringkasan yang dihasilkan oleh model BART.

Ketiga, dalam model Transformer yang dibangun dari awal summarize dilakukan dengan memanfaatkan model T5 (Text-to-Text Transfer Transformer). Fungsi generate_summary dirancang untuk menerima teks input, mengkodekannya menggunakan tokenizer, dan kemudian memasukkannya ke dalam model T5 untuk menghasilkan ringkasan. Dalam proses ini, digunakan teknik beam search untuk menghasilkan ringkasan yang optimal dari teks input. Setelah proses ringkasan selesai, menggunakan pustaka ROUGE, sistem memberikan skor untuk mengukur akurasi ringkasan yang dihasilkan oleh model T5. Dengan demikian, proses ini memberikan gambaran objektif terhadap kualitas ringkasan yang dihasilkan oleh model Transformer dari awal.

2.9. Penentuan Nilai Rouge

Nilai ROUGE (Recall-Oriented Understudy for Gisting Evaluation) digunakan untuk mengevaluasi kualitas sebuah ringkasan yang dihasilkan oleh model, baik itu dibandingkan dengan ringkasan referensi yang telah disediakan maupun dibandingkan dengan ringkasan manusia. Penerapan metode ini dapat memberikan gambaran yang objektif tentang seberapa baik sebuah model melakukan tugas ringkasan teks.

Pertama, dalam model BART penentuan nilai ROUGE dilakukan dengan membandingkan ringkasan yang dihasilkan oleh model dengan ringkasan referensi yang telah tersedia. Metrik ROUGE seperti ROUGE-1, ROUGE-2, dan ROUGE-L digunakan untuk mengukur kesamaan antara ringkasan model dengan ringkasan referensi. Dengan menggunakan berbagai metrik ini, skor ROUGE memberikan indikasi tentang seberapa baik model BART mampu menghasilkan ringkasan yang mirip dengan referensi.

Kedua, model BERT juga menggunakan metode ROUGE untuk mengevaluasi kualitas ringkasan yang dihasilkannya. Setelah model BERT menghasilkan ringkasan dari teks input, ROUGE

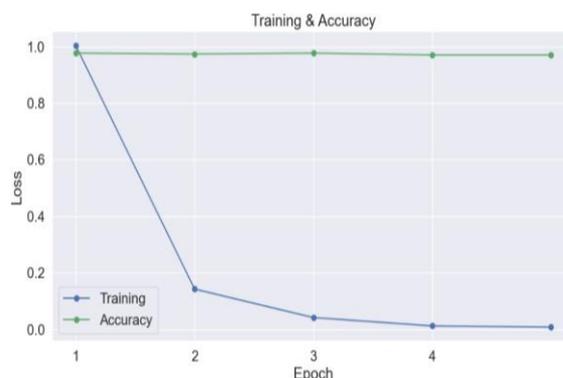
digunakan untuk mengukur kesamaannya dengan ringkasan referensi. Berbagai metrik ROUGE digunakan untuk memberikan gambaran yang lebih lengkap tentang kualitas ringkasan yang dihasilkan oleh model BERT.

Ketiga, proses penentuan nilai ROUGE dalam model Transformer, seperti T5, dilakukan dengan cara yang serupa. Setelah model menghasilkan ringkasan dari teks input, ROUGE digunakan untuk membandingkan kesamaannya dengan ringkasan referensi yang telah disediakan. Dengan menggunakan metrik ROUGE-1, ROUGE-2, dan ROUGE-L, evaluasi dapat dilakukan untuk menilai seberapa baik model Transformer dari awal mampu menangkap inti atau esensi dari teks input. Dengan demikian, metode ROUGE memberikan landasan objektif untuk menilai kualitas ringkasan yang dihasilkan oleh berbagai model transformer.

3. RESULT

Dalam penelitian ini, terdapat tiga model transformer yang digunakan adalah BERT, BART, dan Transformer from scratch. Berikut merupakan hasil dari training dari setiap model tersebut :

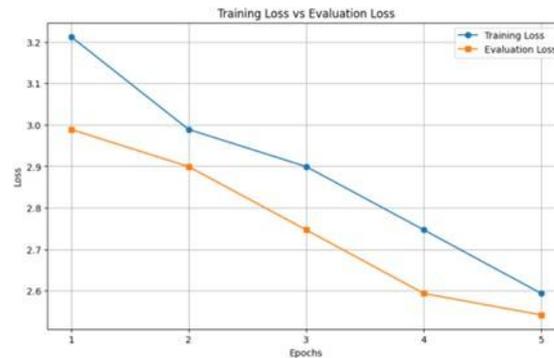
Pertama, Hasil pelatihan dengan model BERT menunjukkan penurunan secara signifikan dalam nilai loss selama lima epoch. Pada epoch pertama, nilai loss pelatihan adalah 1.003, yang kemudian turun menjadi 0.142 pada epoch kedua. Penurunan berlanjut hingga epoch keempat, di mana nilai loss pelatihan mencapai hanya 0.012. Penurunan loss yang signifikan ini menunjukkan bahwa model secara efektif mempelajari pola-pola dalam data pelatihan. Hasil validasi juga menunjukkan kinerja yang baik dari model BERT. Nilai loss validasi turun dari 0.247 pada epoch pertama menjadi 0.081 pada epoch kedua, menunjukkan bahwa model tidak hanya berkinerja baik pada data pelatihan tetapi juga mampu menggeneralisasi dengan baik pada data validasi yang belum pernah dilihat sebelumnya. Selain itu, akurasi validasi juga tetap tinggi sepanjang pelatihan, dengan nilai sekitar 97%. Waktu pelatihan sedikit berfluktuasi dari epoch ke epoch, rata-rata waktu pelatihan sekitar 9 hingga 11 menit per epoch. Demikian pula, waktu validasi relatif singkat, dengan rata-rata sekitar 51 hingga 60 detik per epoch. Hal ini menunjukkan bahwa pelatihan model dilakukan dengan efisien tanpa mengorbankan waktu yang berlebihan. Hasil pelatihan dengan model BERT menunjukkan peningkatan kinerja yang stabil seiring berjalannya waktu. Dengan penurunan nilai loss dan tingkat akurasi yang tinggi, serta waktu pelatihan dan validasi yang efisien, model BERT berhasil melatih dengan baik dan mampu memberikan hasil yang memuaskan. Pada Gambar 5 berikut merupakan grafik dari hasil training mode BERT.



Gambar 5. Hasil Training Model BERT

Kedua, Proses training dengan model BART menunjukkan terdapat penurunan nilai loss evaluasi secara bertahap selama lima epoch pelatihan. Pada awal pelatihan, nilai loss evaluasi adalah sekitar 4.68 dan kemudian turun menjadi sekitar 4.17 pada akhir pelatihan. Penurunan ini menunjukkan bahwa model BART secara bertahap mempelajari pola-pola dalam data pelatihan dan meningkatkan kualitas ringkasan yang dihasilkan. Meskipun terjadi penurunan loss, hasil evaluasi menunjukkan

kecenderungan yang cukup stabil dalam peningkatan kualitas ringkasan. Hal ini dapat dilihat dari peningkatan kinerja evaluasi pada setiap iterasi. Meskipun tidak signifikan, peningkatan ini menunjukkan bahwa model terus mengasah kemampuannya selama proses pelatihan. Waktu pelatihan yang tercatat juga patut diperhatikan. Proses pelatihan memakan waktu sekitar 16 menit dan 28 detik untuk lima epoch, dengan rata-rata sekitar 19.78 detik per iterasi. Meskipun demikian, kecepatan pelatihan tergantung pada konfigurasi perangkat keras dan ukuran dataset yang digunakan. Hasil pelatihan dengan model BART menunjukkan peningkatan yang stabil dalam kualitas ringkasan seiring berjalannya waktu. Meskipun penurunan loss pada evaluasi terjadi secara bertahap, model berhasil memperbaiki kualitas ringkasan dengan efektif selama proses pelatihan. Pada Gambar 6 berikut merupakan grafik dari hasil training mode BART:



Gambar 6. Hasil Training Model BART

Ketiga, Hasil pelatihan dengan model Transformer dari awal menunjukkan penurunan nilai loss secara konsisten seiring berjalannya waktu pada setiap epoch. Pada epoch pertama, nilai loss pelatihan adalah 0.798 yang kemudian menurun secara signifikan menjadi 0.406 pada epoch terakhir. Penurunan ini menunjukkan bahwa model secara efektif mempelajari pola-pola dalam data pelatihan dan semakin meningkatkan kinerjanya seiring dengan iterasi pelatihan. Tingkat akurasi pelatihan juga meningkat seiring berjalannya waktu. Pada awal pelatihan, tingkat akurasi pelatihan sekitar 76%, dan meningkat menjadi sekitar 79% pada akhir pelatihan. Hal ini menunjukkan bahwa model tidak hanya efektif dalam menurunkan loss tetapi juga semakin baik dalam memprediksi label yang benar seiring berjalannya waktu. Waktu pelatihan relatif cepat, dengan rata-rata sekitar 1 menit dan 50 detik per epoch. Meskipun demikian, model berhasil mencapai peningkatan kinerja yang signifikan dalam waktu yang relatif singkat. Selain itu, selama pelatihan, berbagai metrik seperti akurasi dan loss dipantau secara real-time, dan model menyimpan bobot terbaiknya untuk digunakan di masa depan. Hasil pelatihan dengan model Transformer dari awal menunjukkan peningkatan yang signifikan dalam kinerja model seiring dengan berjalannya waktu. Dengan penurunan nilai loss, peningkatan tingkat akurasi, dan waktu pelatihan yang efisien, model berhasil mempelajari pola-pola yang kompleks dalam data dan memberikan hasil yang memuaskan.

3.1. Pembagian Data

Pembagian data dilakukan setelah data yang dikumpulkan dalam proses summarize. Datasheet akan dibagi menjadi 2 bagian yaitu untuk proses training dan proses testing dengan persentase sebagai berikut. Datasheet akan dibagi menjadi 2 bagian yaitu untuk proses training dan proses testing dengan persentase yang disajikan dalam tabel 2.

Tabel 2. Persentase Pembagian Data

No	Pembagian Data	Presentase
----	----------------	------------

1.	Data Training	80%
2.	Data Testing	20%

3.2. Tokenisasi

Dalam penelitian ini dilakukan proses tokenisasi untuk mempersiapkan teks untuk diproses oleh model-model BERT, BART, dan Transformer, tokenisasi dilakukan dengan cara yang spesifik sesuai dengan arsitektur dan kebutuhan model tersebut.

Hasil dari proses tokenisasi adalah urutan token-token yang telah diproses dan siap untuk digunakan dalam `[CLS]` di awal dan `[SEP]` di akhir setiap kalimat. Ini membantu model dalam memahami struktur teks. Selain itu, jumlah token dalam setiap kalimat akan disesuaikan agar mencapai panjang maksimum yang telah ditentukan. Hasil tokenisasi, berupa input_ids yang berisi ID-ID token dan attention masks yang membedakan antara kata-kata asli dan padding, siap untuk digunakan dalam proses training atau inference.

Kedua, pada model BART, proses tokenisasi dilakukan menggunakan auto tokenizer yang sesuai dengan model tersebut. Teks dimasukkan ke dalam tokenizer dengan penanganan padding dan truncation sesuai dengan panjang maksimum yang ditentukan. Hasil tokenisasi berupa encodings, yang berisi token-token yang telah diproses dan siap digunakan oleh model BART. Ada juga penambahan <s> di awal kalimat dan </s> diakhir kalimat.

Ketiga, Model Transformer yang dibangun dari awal juga melibatkan proses tokenisasi yang khusus. Teks dipecah menjadi token menggunakan tokenizer yang sesuai, dan setiap kata diubah menjadi huruf kecil. Untuk membantu model memahami dokumen awal dan akhir teks saat merangkum dokumen, token `_START_` ditambahkan di awal teks dan `_END_` di akhir teks. Hasil dari proses tokenisasi adalah urutan token-token yang telah diproses dan siap untuk digunakan dalam model Transformer. Hasil tokenisasi pada setiap model disajikan dalam tabel 3 berikut:

Tabel 3. Hasil Tokenisasi pada Setiap Model

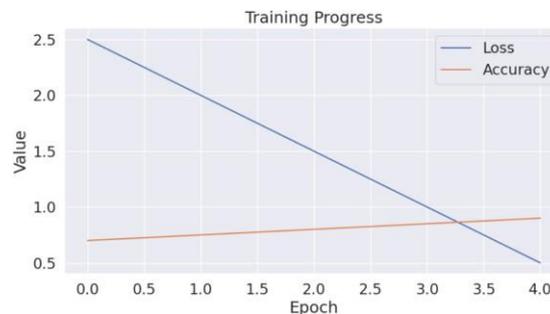
No	Category	Text	Tokenisasi BERT	Tokenisasi BART	Tokenisasi Transformer From Scratch
1.	tech	tv future in the hands of viewers with home theatre systems ...	[[CLS]', 'tv', 'future', 'in', 'the', 'hands', 'of', 'viewers', 'with', 'home', 'theatre', 'systems', '...', '[SEP]']	<s>', 'sou', ##ness', 'delight', 'at', 'euro', 'progress', '...', </s>	['START', 'tv', 'future', 'in', 'the', 'hands', 'of', 'viewers', 'with', 'home', 'theatre', 'systems', '...', 'END']
2.	business	worldcom boss left books alone former worldcom ...	[[CLS]', 'world', ##com', 'boss', 'left', 'books', 'alone', 'former', 'world', ##com', '...', '[SEP]']	<s>', 'world', ##com', 'boss', 'left', 'books', 'alone', 'former', 'world', ##com', '...', </s>	['START', 'world', ##com', 'boss', 'left', 'books', 'alone', 'former', 'world', ##com', '...', 'END']
...
2226.	sport	souness delight at euro progress ...	[[CLS]', 'sou', ##ness', 'delight', 'at', 'euro', 'progress', '...', '[SEP]']	<s>', 'sou', ##ness', 'delight', 'at', 'euro', 'progress', '...', </s>	['START', 'sou', ##ness', 'delight', 'at', 'euro', 'progress', '...', 'END']

'progress', '...', 'progress', '...',
[SEP] </s>

3.3. Training Model

Dalam penelitian ini, terdapat tiga model transformer yang digunakan adalah BERT, BART, dan Transformer from scratch. Berikut merupakan hasil dari training dari setiap model tersebut :

Pertama, Hasil pelatihan dengan model BERT menunjukkan penurunan secara signifikan dalam nilai kerugian (loss) selama lima epoch. Pada epoch pertama, nilai kerugian pelatihan adalah 1.003 yang kemudian turun menjadi 0.142 pada epoch kedua. Penurunan berlanjut hingga epoch keempat, di mana nilai kerugian pelatihan mencapai hanya 0.012. Penurunan kerugian yang signifikan ini menunjukkan bahwa model secara efektif mempelajari pola-pola dalam data pelatihan. Hasil validasi juga menunjukkan kinerja yang baik dari model BERT. Nilai kerugian validasi turun dari 0.247 pada epoch pertama menjadi 0.081 pada epoch kedua, menunjukkan bahwa model tidak hanya berkinerja baik pada data pelatihan tetapi juga mampu menggeneralisasi dengan baik pada data validasi yang belum pernah dilihat sebelumnya. Selain itu, akurasi validasi juga tetap tinggi sepanjang pelatihan, dengan nilai sekitar 97%. Waktu pelatihan sedikit berfluktuasi dari epoch ke epoch, rata-rata waktu pelatihan sekitar 9 hingga 11 menit per epoch. Demikian pula, waktu validasi relatif singkat, dengan rata-rata sekitar 51 hingga 60 detik per epoch. Hal ini menunjukkan bahwa pelatihan model dilakukan dengan efisien tanpa mengorbankan waktu yang berlebihan. Hasil pelatihan dengan model BERT menunjukkan peningkatan kinerja yang stabil seiring berjalannya waktu. Dengan penurunan nilai kerugian dan tingkat akurasi yang tinggi, serta waktu pelatihan dan validasi yang efisien, model BERT berhasil melatih dengan baik dan mampu memberikan hasil yang memuaskan. Pada gambar 7 berikut merupakan grafik dari hasil training mode Transformer:



Gambar 7. Hasil Training Model Transformer From Scratch

Dari hasil training pada ketiga model tersebut maka, akan dilanjutkan pada tahap summarize dokumen. Berikut merupakan hasil dalam rangkuman dokumen dalam Bahasa Inggris pada index ke 0 atau dokumen pertama dari setiap model tersebut. Pada Tabel 3 merupakan Hasil Summarize Dokumen dengan Model BERT:

Tabel 3. Hasil Summarize Dokumen dengan Model BART

No	Category	Text	Summary Asli	Summary Sistem
1.	tech	tv future in the hands of viewers with home theatre systems plasma high-definition tvs and digital video recorders moving into the living	what this means said stacey jolna senior vice president of tv guide tv group is that the way people find the content they want	Future TV viewing shifts to personalized experiences with technologies like DVRs, high-definition TVs, and home networks. The US leads in these innovations, allowing viewers

room the way people watch tv will be radically different in five years time ...	to watch has to be simplified for tv viewers ...	to customize content delivery and enjoy a-la-carte entertainment ...
--	--	--

Tabel 4. Hasil Summarize Dokumen dengan Model BERT

No	Category	Text	Summary Asli	Summary Sistem
1.	tech	tv future in the hands of viewers with home theatre systems plasma high-definition tvs and digital video recorders moving into the living room the way people watch tv will be radically different in five years time ...	what this means said stacey jolna senior vice president of tv guide tv group is that the way people find the content they want to watch has to be simplified for tv viewers ...	tv future in the hands of viewers with home theatre systems plasma high - definition tvs and digital video recorders moving into the living room the way people watch tv will be radically different in five years time ...

Langkah selanjutnya adalah penentuan nilai rouge untuk mengetahui seberapa akurat model dalam melakukan peringkasan dokumen berita. Berikut merupakan Hasil Penentuan Nilai Rouge dalam rangkuman dokumen dalam Bahasa inggris pada index ke 0 atau dokumen pertama dari setiap model. Pada Tabel 7, 8, 9, dan 10 merupakan hasil testing pada setiap model:

Tabel 5. Hasil Nilai Rouge Model BERT

No	Category	Text	Rouge-1	Rouge-2	Rouge-L
1.	tech	tv future in the hands of viewers with home theatre systems plasma high ...	0.247	0.159	0.159

Tabel 6. Hasil Nilai Rouge Model BART

No	Category	Text	Rouge-1	Rouge-2	Rouge-L
1.	tech	tv future in the hands of viewers with home theatre systems plasma high ...	0.524	0.352	0.369

Tabel 7. Hasil Nilai Rouge Model Transformer From Scratch

No	Category	Text	Rouge-1	Rouge-2	Rouge-L
1.	tech	tv future in the hands of viewers with home theatre systems plasma high ...	0.702	0.565	0.638

Tabel 8. Perbandingan Hasil Nilai Rouge

No	Model	Rouge-1	Rouge-2	Rouge-L
1	BART	0.523	0.523	0.368
2	BERT	0.247	0.159	0.159
3	Transformer Form Scratch	0.702	0.565	0.638

4. DISCUSSIONS

Perbedaan kinerja yang signifikan antara model Transformer from Scratch, BART, dan BERT dalam tugas peringkasan dokumen. Secara keseluruhan, Transformer from Scratch berhasil menghasilkan skor tertinggi dalam beberapa kategori evaluasi ROUGE, sedangkan BART dan BERT menunjukkan keunggulan dalam aspek yang berbeda. Diskusi ini akan menghubungkan hasil tersebut

dengan penelitian terdahulu serta memberikan penjelasan mengenai faktor-faktor yang mungkin menyebabkan perbedaan kinerja antara ketiga model tersebut.

Salah satu temuan adalah kinerja unggul Transformer from Scratch dalam beberapa matrik evaluasi ROUGE, yang mengindikasikan bahwa model ini, meskipun memerlukan pelatihan lebih lama dan data yang lebih besar, dapat disesuaikan secara spesifik untuk tugas tertentu, seperti peringkasan dokumen. Keunggulan ini selaras dengan penelitian sebelumnya yang menunjukkan bahwa model transformer yang dibangun dari awal dapat memberikan fleksibilitas lebih besar dalam penyesuaian arsitektur untuk tugas-tugas spesifik. Dengan menyesuaikan jumlah lapisan, ukuran hidden layer, dan attention heads, model ini mampu menangkap pola yang lebih kompleks dalam teks, memberikan hasil yang lebih presisi.

BART, di sisi lain, menunjukkan kinerja yang kuat dalam menghasilkan ringkasan yang koheren dan terstruktur. Kemampuannya dalam melakukan denoising pada input teks dan penggunaan pendekatan autoregressive memberikan keuntungan dalam menghasilkan ringkasan yang berkualitas tinggi, terutama dalam teks yang memerlukan pengolahan konteks yang lebih dalam. Penelitian sebelumnya juga mendukung keunggulan BART dalam tugas-tugas seperti generasi teks dan summarization, di mana model ini terbukti lebih efisien dalam menangani teks yang panjang dan rumit.

BERT, yang dikenal dengan kekuatan dalam memahami konteks dua arah, memberikan hasil yang baik, namun tidak seunggul Transformer from Scratch dalam beberapa kategori. Ini mungkin disebabkan oleh fakta bahwa BERT awalnya dirancang untuk tugas klasifikasi teks dan tidak secara khusus dioptimalkan untuk tugas generasi teks seperti summarization. Penelitian terdahulu juga menyebutkan bahwa meskipun BERT mampu menangkap konteks yang lebih kaya, penggunaannya dalam tugas-tugas peringkasan memerlukan penyesuaian.

Dengan demikian, penelitian ini memberikan kelebihan dan kekurangan masing-masing model dalam peringkasan dokumen, serta mendukung temuan dari penelitian terdahulu mengenai potensi besar model transformer dalam aplikasi Natural Language Processing (NLP).

5. CONCLUSION

Penelitian ini mengungkapkan bahwa ketiga model Transformer-BERT, BART, dan Transformer from Scratch-memiliki kelebihan dan kekurangannya masing-masing dalam tugas peringkasan dokumen. BERT unggul dalam memahami konteks dua arah teks dan memberikan hasil yang menjanjikan dalam meringkas dokumen yang kompleks. BART, dengan kemampuan denoise dan output yang koheren, menunjukkan kinerja yang kuat. Model Transformer from Scratch, meskipun membutuhkan data pelatihan yang besar dan sumber daya komputasi yang signifikan, mampu disesuaikan dengan kebutuhan spesifik dan memberikan hasil yang unggul dalam peringkasan dokumen jika dilatih dengan benar.

Model BERT menunjukkan penurunan nilai loss secara signifikan dari 1.003 pada epoch pertama menjadi 0.012 pada epoch keempat, dengan akurasi validasi yang tinggi sekitar 97% sepanjang pelatihan. Model BART juga menunjukkan peningkatan kualitas ringkasan dengan penurunan nilai loss dari 4.68 menjadi 4.17 selama lima epoch pelatihan. Sementara itu, model Transformer from Scratch, meskipun memerlukan data pelatihan yang besar dan sumber daya komputasi yang signifikan, berhasil memberikan hasil yang unggul dengan nilai ROUGE-1 sebesar 0.7021, ROUGE-2 sebesar 0.5652, dan ROUGE-L sebesar 0.6383. Pemilihan model peringkasan dokumen berbasis Transformer harus mempertimbangkan kebutuhan spesifik tugas dan sumber daya yang tersedia. Kombinasi pendekatan yang sudah terlatih seperti BERT dan BART dengan model yang dibuat dari awal seperti Transformer from Scratch dapat menghasilkan solusi yang lebih efektif dan efisien dalam berbagai konteks aplikasi NLP.

REFERENCES

- [1] B. A. K. Balouch and F. Hussain, "A Transformer based approach for Abstractive Text Summarization of Radiology Reports," *International Conference on Applied Engineering and Natural Sciences*, vol. 1, no. 1, pp. 476–486, 2023, doi: 10.59287/icaens.1042.
- [2] H. Lyu, N. Sha, S. Qin, M. Yan, Y. Xie, and R. Wang, "Manifold Denoising by Nonlinear Robust Principal Component Analysis," *Adv Neural Inf Process Syst*, vol. 32, no. NeurIPS, pp. 1–11, 2019, [Online]. Available: <http://arxiv.org/abs/1911.03831>
- [3] M. Lewis *et al.*, "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 7871–7880, 2020, doi: 10.18653/v1/2020.acl-main.703.
- [4] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, no. M1m, pp. 4171–4186, 2019.
- [5] C. Raffel *et al.*, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *Journal of Machine Learning Research*, vol. 21, pp. 1–67, 2020, Accessed: Jul. 24, 2024. [Online]. Available: <http://jmlr.org/papers/v21/20-074.html>.
- [6] I Nyoman Purnama and Ni Nengah Widya Utami, "IMPLEMENTASI PERINGKAS DOKUMEN BERBAHASA INDONESIA MENGGUNAKAN METODE TEXT TO TEXT TRANSFER TRANSFORMER (T5)," *Jurnal Teknologi Informasi dan Komputer*, vol. 9, no. 4, Aug. 2023, doi: 10.36002/jutik.v9i4.2531.
- [7] F. Raihanunnisa, M. Arhami, and R. Hidayat, "PENDEKATAN HYBRID PADA SISTEM PERINGKAS TEKS ARTIKEL BERITA BAHASA INGGRIS MENGGUNAKAN NATURAL LANGUAGE PROCESSING", Accessed: Jul. 24, 2024. [Online]. Available: <https://journal.budiluhur.ac.id/index.php/telematika/>
- [8] A. JAISWAL, S. Liu, T. Chen, and Z. "Atlas" Wang, "The Emergence of Essential Sparsity in Large Pre-trained Models: The Weights that Matter," *Adv Neural Inf Process Syst*, vol. 36, pp. 38887–38901, Dec. 2023, Accessed: Jul. 24, 2024. [Online]. Available: https://github.com/VITA-Group/essential_sparsity.
- [9] F. V. P. Samosir, H. Toba, and M. Ayub, "BESKlus : BERT Extractive Summarization with K-Means Clustering in Scientific Paper," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 1, 2022, doi: 10.28932/jutisi.v8i1.4474.
- [10] A. A. Fattahila, A. Romadhony, and S. Al Faraby, "Peringkasan Artikel Berita Menggunakan Pendekatan Abstraktif Dengan Model Transformers," vol. 10, no. 5, pp. 4980–4986, 2023.
- [11] Mulaab, M. (2020). RINGKASAN DOKUMEN ILMIAH BERBASIS REPRESENTASI DATA GRAPH MENGGUNAKAN UKURAN CENTRALITY. *Jurnal Simantec*, 9(1), 33-38.
- [12] F. Noprianto, S. Agustian, M. Irsyad, N. Sultan, and S. K. Riau, "Clustering Peringkasan Teks Otomatis Dokumen Berita menggunakan Metode K-Means," pp. 139–147, 2023, [Online]. Available: <https://prosiding.unipma.ac.id/index.php/sendiko/article/view/3911/3569>
- [13] S. HENI, "Implementasi Berbagai Metode Kecerdasan Buatan (Artificial Intelligence) Pada Masalah Gangguan Kepribadian (Narcissistic ...)," *Digital Repostiry UNILA*, 2023, [Online]. Available: <http://digilib.unila.ac.id/78126/>
- [14] Rahman, S., Sembiring, A., Aulia, R., Dafitri, H., & Liza, R. (2023). Pengenalan ChatGPT untuk Meningkatkan Pengetahuan Siswa-Siswi di SMK Negeri 1 Pantai Labu. *Prioritas: Jurnal Pengabdian Kepada Masyarakat*, 5(01), 1-7.
- [15] Yuliska, Y., & Syaliman, K. U. (2020). Literatur Review Terhadap Metode, Aplikasi dan Dataset Peringkasan Dokumen Teks Otomatis untuk Teks Berbahasa Indonesia. *IT Journal Research and Development*, 5(1), 19-31.
- [16] Al. Nasir, et, "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title," vol. 9, pp. 356–363, 2023.

-
- [17] N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, and R. Socher, "CTRL: A Conditional Transformer Language Model for Controllable Generation," pp. 1–18, 2019, [Online]. Available: <http://arxiv.org/abs/1909.05858>
- [18] Q. Fan, H. Huang, X. Zhou, and R. He, "Lightweight Vision Transformer with Bidirectional Interaction," *Adv Neural Inf Process Syst*, vol. 36, no. NeurIPS, 2023.
- [19] B. Juarto and Yulianto, "Indonesian News Classification Using IndoBert," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 2, pp. 454–460, 2023.
- [20] B. Kurniawan, A. Ari Aldino, and A. Rahman Isnain, "Sentimen Analisis Terhadap Kebijakan Penyelenggara Sistem Elektronik (PSE) Menggunakan Algoritma Bidirectional Encoder Representations From Transformer (BERT)," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 3, no. 4, pp. 98–106, 2022, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [21] H. E. Syah, D. F. Huwaida, and L. Danuarta, "Efisiensi Manajemen Sekolah Dan Mutu Pembelajaran," vol. 21, no. 1, pp. 15–22, 2021.
- [22] A. Alokla, W. Gad, W. Nazih, M. Aref, and A. B. Salem, "Pseudocode Generation from Source Code Using the BART Model," *Mathematics*, vol. 10, no. 21, 2022, doi: 10.3390/math10213967.
- [23] S. Nabilah, "Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Novel Online di Media Sosial Menggunakan Latent Dirichlet Allocation dan Bidirectional Encoder Representation from Transformers," 2022, [Online]. Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/65262>
- [24] S. McDonnell, O. Nada, M. R. Abid, and E. Amjadian, "CyberBERT: A Deep Dynamic-State Session-Based Recommender System for Cyber Threat Recognition," *IEEE Aerospace Conference Proceedings*, vol. 2021-March, no. March, 2021, doi: 10.1109/AERO50100.2021.9438286.
- [25] D. Setiawan, E. A. D. Karuniawati, and S. I. Janty, "Peran Chat Gpt (Generative Pre-Training Transformer) Dalam Implementasi Ditinjau Dari Dataset," *INNOVATIVE: Journal of Social Science Research*, vol. 3, no. 3, pp. 9527–9539, 2023, [Online]. Available: <https://j-innovative.org/index.php/Innovative/article/view/3286>
- [26] M. T. Anwar, L. Heriyanto, and F. Fanini, "Model Prediksi Dropout Mahasiswa Menggunakan Teknik Data Mining," *Jurnal Informatika Upgris*, vol. 7, no. 1, pp. 56–60, 2021, doi: 10.26877/jiu.v7i1.8023.
- [27] A. Jaiswal, S. Liu, T. Chen, and Z. Wang, "The Emergence of Essential Sparsity in Large Pre-trained Models: The Weights that Matter," *Adv Neural Inf Process Syst*, vol. 36, no. NeurIPS, pp. 1–15, 2023.
- [28] Farah Raihanunnisa, Muhammad Arhami, and Rahmad Hidayat, "Pendekatan Hybrid Pada Sistem Peringkat Teks Artikel Berita Bahasa Inggris Menggunakan Natural Language Processing," *Telematika Mkom*, vol. 15, no. 2, pp. 86–92, 2023, [Online]. Available: <https://journal.budiluhur.ac.id/index.php/telematika/article/view/2679>
-

